Effective Metadata Design

Michael C. Daconta

Chief, Enterprise Data Management

Oberon Associates, Inc. and

Former Metadata Program Manager, DHS

mdaconta@oberonassociates.com

http://www.daconta.us

Author:
Current Position:
Company:
Previous Position:
Email:
Website:

Agenda

- The best example of Metadata

- Explaining Metadata to your boss

- Redefining Metadata

- Metadata Types, Targets and Tao

- Methods of Description

- Applying these Techniques

- Conclusion

The BEST example of Metadata

- **Test**: Would you take a music player full of unmarked songs and use them for your wedding?

- Music Data
  - MP3, WMA, AAC… etc.
  - Data Format (aka structure)
    - Version, Layer, Bit Rate, Frequency, etc.

- Music Metadata
  - Artist, Year, Genre, Album, Song Title, Duration, etc.

- Experimental
  - Mood (Moody)
  - Similarity (Pandora)
  - Cover Flow (Apple)

Now Playing

1 of 14

Dani California
Red Hot Chili Peppers
Stadium Arcadium

3:27   -1:15
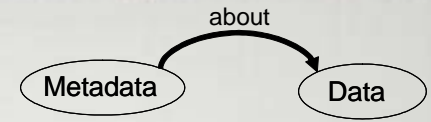
MENU

© Apple

Explaining Metadata to your Boss

- Product Analogy

- Packaging

- Labeling

- Advertising

- Store Placement

- Yellow Pages Directory
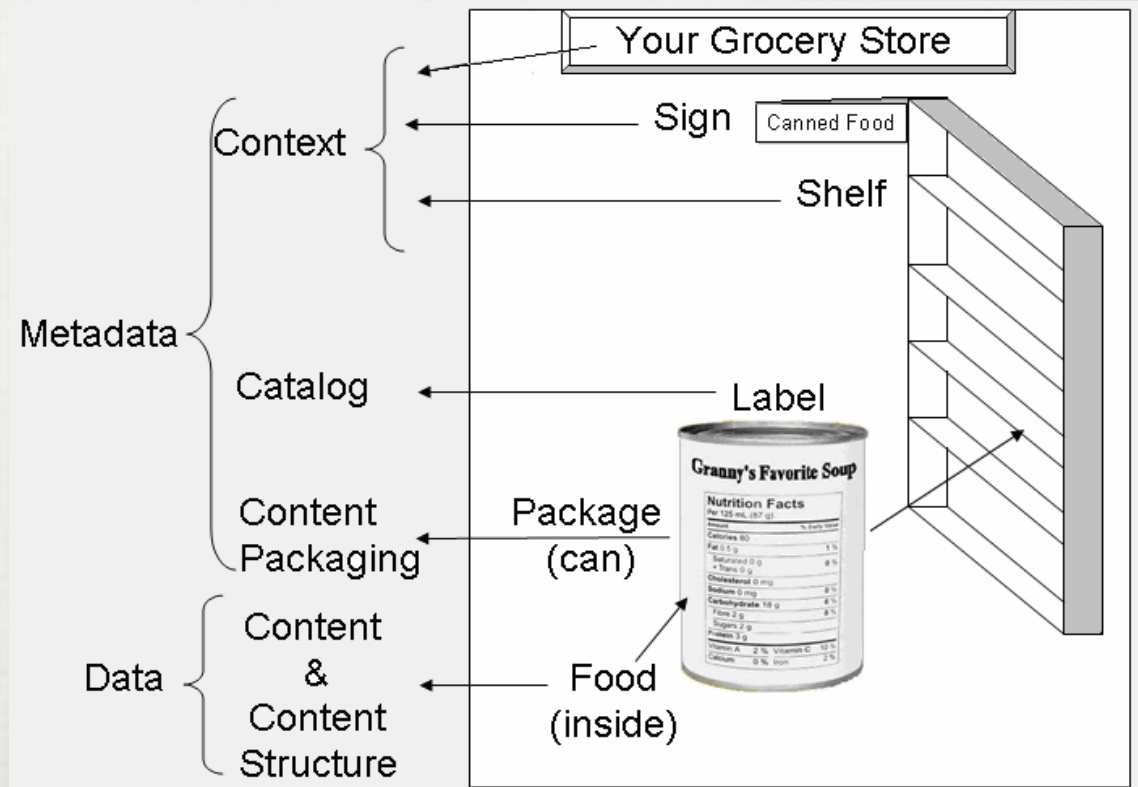
© 2007 Michael C. Daconta

Redefining Metadata

- Classical definition considered harmful!

- Understanding "aboutness"…

- Countering naysayers …

© 1997 by LBNL

> **Metadata** – an external description of a distinct data resource. Common usages for metadata include providing the context of the data resource, managing its lifecycle and extending it to new uses. An example of metadata is the external description of an audio file specifying the artist that created it, when it was created, the length or play time, and the genre of music it belongs to. The purpose of metadata is to manage and improve the use of data thereby turning it into a strategic information asset.

IPTC Metadata
- Date Created :
- Intellectual Genre:
- IPTC Scene:
- Location
- City
- State/Province …

Metadata Types, Targets and Tao

- Metadata Types
  - Roles/Purpose: General Context, Lineage, Performance Metrics

- Metadata Targets
  - Wrong focus, i.e. technical, business, etc. … targets will continue to multiply exponentially.

- The Tao of Metadata –
  - The yin of metadata is description.
  - The yang of metadata is use.
  - Think of a management layer above the data …

Methods of Description

- Identification

- Static Measurement

- Dynamic Measurement

- Degree

- Categorization
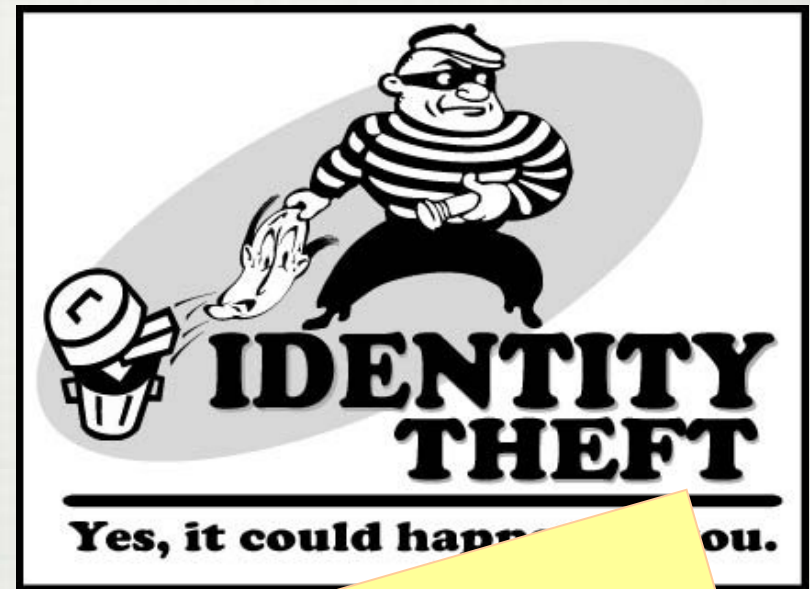
- Relationships

- Commentary

© MGM

Identification

Identification represents the ability to distinguish one data asset from another.

- Unique?
  - GUID, UUID, URI, etc.

- Dereference-able?
  - URL

- Semantics?
  - UDEF, Semantic URIs, etc.

- Privacy Implications?
  - Non-SSN IDs

- Example: How do you uniquely identify a song? (CDDB)

Identification
- **Name:**
- **<entity>ID:**
- **Location:**
- **Signatures:**

Static Measurement

This is a measure of relatively constant characteristics of a target data asset.

- Metadata measures externals…

- How to measure fixed characteristics:
  - Observables
  - Placement in an Environment
  - Effect

- Example:
  - What are the fixed characteristics of a Song?
    - Format, Length…
    - Creation Date, Creator, etc.

Static Measures

- Format:
- Physical Size:
- Creator:
- Security
- Content-Specific:

9

Dynamic Measurement

This technique measures variable or changing aspects of a data asset.

- How to measure dynamic characteristics:
  - Tracking
    - Movement through time & space
  - Pace
  - Intensity

- Example:
  - What are the dynamic aspects of a Song?
    - Ratings
    - Sales
    - Plays
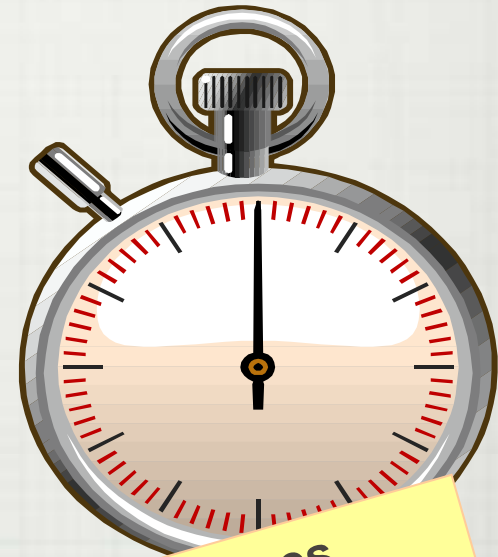
Dynamic Measures

- Usage Counts:
- <negative>Counts:
- Location:
- Ranking:

Degree

Degree is the scale by which an artifact may be measured along a continuum and the meaningful inflection points along that continuum.

- First define the endpoints
    - Define the inflection points

- Align granularity with measurement ability

- Very useful in scorecards

- Example:
    - What type of scales have been designed for Music?
        - Music charts (top 40)
        - Mood, Tempo

© 2008 by Michael C. Daconta

Degree
- Time scales:
- Performance scales:
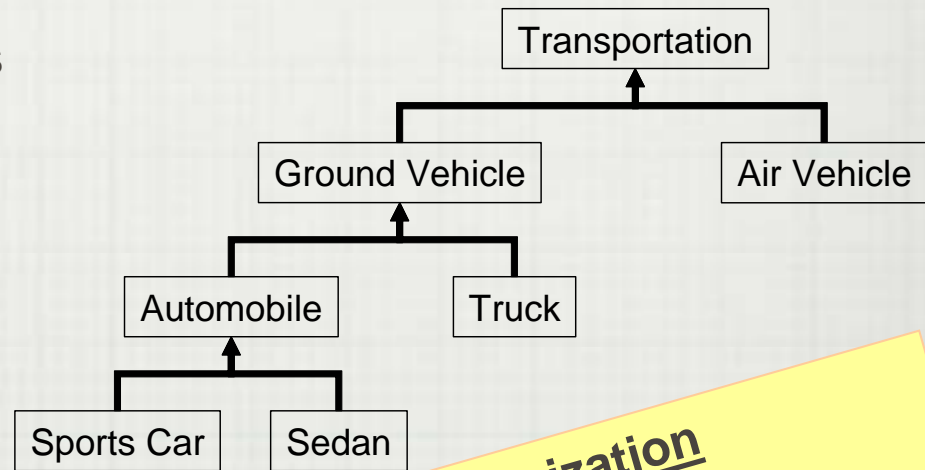- Opinion scales:
- Imbuement:

Categorization

This technique enables the division of a population of data assets into manageable groups based on commonalities of all the members within a group.

- A good taxonomy has a theme
  - The theme binds the elements
  - TSA example

- Formal Taxonomies
  - www.xml.com article
  - Formal Nodes:
    - Class, Instance, Collection
  - Formal Relationships
    - Subclass of, Part of, Instance Of
  - Example:
    - What is the categorization scheme for music? (Genre/sub-genre)

```
                        Transportation
                         ↑
            Ground Vehicle        Air Vehicle
                 ↑
       Automobile      Truck
           ↑
   Sports Car    Sedan
```

Categorization

- **Class:**
- **Instance:**
- **Subclass:**
- **PartOf:**

Relationships

This technique creates predicates (also known as relationships) between the metadata record and its target data asset

- Advanced Technique

- Social Networks
  - Friend of a Friend (FOAF)

- Critical to implement Rules
  - Web Ontology Language (OWL)

- Example:
  - What relationships exist between music data?
    - Cover Art, Band Member cross-pollination, Multimedia connections

Who are YOU connected to?

Relationships
- Symmetric:
- Transitive:
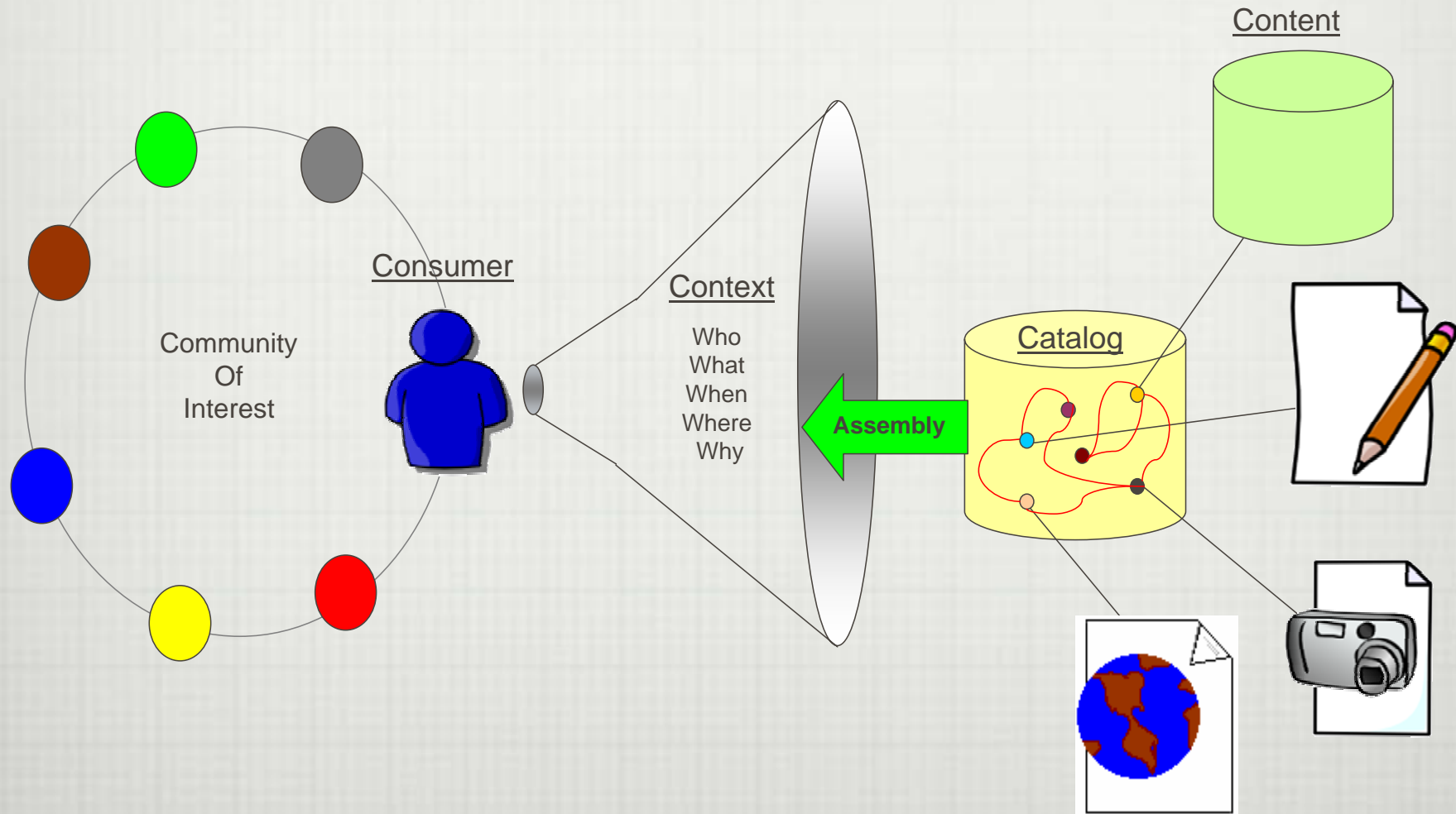- Inverse:

Commentary

This provides free form textual description for human readers of the metadata record.

- Least reliable form of metadata

- Free-form Description

- Full-Text Indexing

- Example:
  - What types of commentary is used for music?
    - Album Reviews
    - Lyrics

Commentary
- Description:
- Incident reports:
- Other:

Applying These Techniques

Content

Consumer

Context

Who
What
When
Where
Why

Community
Of
Interest

Assembly

Catalog

Conclusion

Effective Metadata Design is the key skill in delivering
the Right information to the Right Person at the Right Time

- Effective metadata is an important and necessary step in the Evolution of Data

- We Redefined Metadata

- We explored the Magnificent Seven

- This is the path to effective Information Management and Reliable Information Production!

- Any final Questions?